

## 250 Million Years of Bindin Evolution

KIRK S. ZIGLER<sup>1,2,\*</sup> AND H. A. LESSIOS<sup>1</sup>

<sup>1</sup> *Smithsonian Tropical Research Institute, Balboa, Panamá; and* <sup>2</sup> *Department of Biology, Duke University, Durham, North Carolina*

**Abstract.** Bindin plays a central role in sperm-egg attachment and fusion in sea urchins (echinoids). Previous studies determined the DNA sequence of bindin in two orders of the class Echinoidea, representing 10% of all echinoid species. We report sequences of mature bindin from five additional genera, representing four new orders, including the distantly related sand dollars, heart urchins, and pencil urchins. The six orders in which bindin is now known include 70% of all echinoids, and indicate that bindin was present in the common ancestor of all extant sea urchins more than 250 million years ago. Over this span of evolutionary time there has been (1) remarkable conservation in the core region of bindin, particularly in a stretch of 29 amino acids that has not changed at all; (2) conservation of a motif of basic amino acids at the cleavage site between preprobindin and mature bindin; (3) more than a twofold change in length of mature bindin; and (4) emergence of high variation in the sequences outside the core, including the insertion of glycine-rich repeats in the bindins of some orders, but not others.

### Introduction

Various studies have shown that molecules involved in reproduction (and particularly in gamete interactions) evolve rapidly, often under the influence of positive selection (reviewed in Swanson and Vacquier, 2002). Among these proteins there are examples of both high (Metz and Palumbi, 1996) and low (Metz *et al.*, 1998b) levels of intraspecific variation. In some cases a single molecule displays domains that are highly conserved and other domains that are highly variable (Vacquier *et al.*, 1995). Variation in such proteins is usually studied at a low taxonomic

level, often within species, sometimes within genera, but rarely across an entire class. There are good reasons for this focus: such studies are likely to uncover mutational changes that are important in mate recognition and in speciation. However, comparisons across broad taxonomic levels can offer insights into the evolution of such molecules. They can reveal which features of these molecules are conserved (and are thus essential for basic functions) and which features are free to vary. For the parts that do vary, such comparisons can determine common features of evolution. Most of all, the comparisons can address the question of the universality of a particular molecule by asking how far back in evolution one needs to search to find the point at which a completely different molecule has taken over the essential functions involved in gamete binding and fusion.

Echinoids (sea urchins, heart urchins, and sand dollars), with their readily obtainable gametes, have long been model organisms for fertilization studies. Because fertilization is external, the molecules involved in gamete recognition and fusion are associated exclusively with the gametes. Biochemical studies in sea urchins identified the first “gamete recognition protein,” bindin (Vacquier and Moy, 1977). Bindin is the major insoluble component of the sperm acrosomal vesicle and has been implicated in three molecular interactions (Hofmann and Glabe, 1994). First, after the acrosomal reaction, bindin self-associates, coating the acrosomal process. Second, it functions in sperm-egg attachment by binding to carbohydrates in the vitelline layer on the egg surface. Third, it is involved in the fusion of sperm and egg membranes (Ulrich *et al.*, 1998, 1999).

Bindin is translated as a larger precursor, from which the N-terminal preprobindin portion is subsequently cleaved to produce mature bindin (Gao *et al.*, 1986). The mature bindin molecule contains an amino acid core of about 55 residues that is highly conserved among all bindins characterized to date (Vacquier *et al.*, 1995). An 18 amino acid section of this conserved core (B18) has been shown to fuse lipid

Received 25 February 2003; accepted 3 June 2003.

\* To whom correspondence should be addressed. Current address: Friday Harbor Laboratories, University of Washington, 620 University Road, Friday Harbor, WA 98250. E-mail: ziglerk@u.washington.edu

vesicles *in vitro*, suggesting that this region functions in sperm-egg membrane fusion (Ulrich *et al.*, 1998, 1999). Thus far, bindin is known only from echinoids; no homologous molecules have been identified in any other organism (Vacquier, 1998).

To date, the nucleotide sequence of bindin has been determined in six genera of sea urchins. In *Echinometra* (Metz and Palumbi, 1996), *Strongylocentrotus* (Gao *et al.*, 1986; Minor *et al.*, 1991; Biermann, 1998; Debenham *et al.*, 2000), and *Heliocidaris* (Zigler *et al.*, 2003), there are many sequence rearrangements among individuals and species, and indications of positive selection in regions on either side of the core. In *Arbacia* (Glabbe and Clark, 1991; Metz *et al.*, 1998a) and *Tripneustes* (Zigler and Lessios, 2003), there are fewer sequence rearrangements and no evidence for positive selection. In *Lytechinus*, only one sequence has been published (Minor *et al.*, 1991), so the mode of evolution of the molecule remains unknown.

The five genera in which bindin was previously sequenced belong to two echinoid orders, the Echinoida and the Arbacioida. These two orders contain only 10% of all extant echinoid species (Kier, 1977; Smith, 1984; Littlewood and Smith, 1995). The molecular structure of bindin in the other 13 orders of the class Echinoida has not been studied. The only evidence that bindin is present outside the Echinoida and Arbacioida comes from Moy and Vacquier (1979), who reported that an antibody to bindin of *Strongylocentrotus purpuratus* reacted with sperm from one species of the order Phymosomatoida and two species of the order Clypeasteroida. As Vacquier (1998) has pointed out, molecules that mediate fertilization—in contrast to those central to other basic life processes—often differ between taxa. For example, in the molluscan class Bivalvia, completely different proteins are involved in gamete recognition of oysters (Brandriff *et al.*, 1978) and of mussels (Takagi *et al.*, 1994). It is, therefore, not safe to assume without empirical evidence that bindin is present in all orders of echinoids, or that it has the same general structure as in the taxa in which it has already been characterized.

As a first step in determining which orders of echinoids possess bindin and, if they do, how its structure varies, we cloned and sequenced mature bindin from five genera of sea urchins, four of which belong to orders in which bindin was previously unknown. We combined our data with those of previous studies of bindin in genera belonging to the orders Echinoida and Arbacioida. The final data set includes bindin from 10 genera of sea urchins, pencil urchins, sand dollars, and heart urchins, and the results indicate that the molecule was present in the common ancestor of all extant echinoids that diverged from each other over 250 million years ago. The core sequence has remained remarkably unchanged over this period of time, whereas the areas flanking the core have undergone substantial modification, resulting in great

differences in molecular size, amino acid sequence, and number of repeats.

## Materials and Methods

### Samples

The pencil urchins (order Cidaroida) were represented in our study by *Eucidaris tribuloides*, collected on the Atlantic coast of Panamá; the order Diadematoidea by *Diadema antillarum*, also from the Atlantic coast of Panamá. The sand dollars (order Clypeasteroida) were represented by *Encope stokesii* from the Pacific coast of Panamá; the heart urchins (order Spatangoida) by *Moira clotho* collected at the Perlas Islands in the Bay of Panamá. *Heliocidaris erythrogramma* (order Echinoida) was collected near Sydney, Australia.

### DNA isolation and sequencing

We injected various individuals of each species with 0.5 M KCl until we encountered one that produced sperm. The testes of this ripe male were removed and used either directly for mRNA extraction, or after preservation in either RNALater (Ambion Inc.) or in liquid nitrogen. The methods for mRNA isolation, reverse transcription reactions, initial polymerase chain reactions, 3' and 5' rapid amplification of cDNA ends (RACE) reactions, and DNA sequencing were as described in Zigler and Lessios (2003), with the following modifications. (1) A fragment of the core region of bindin was amplified from the reverse transcriptase reaction product or from genomic DNA, using primers MB1130+ (5'-TGCTSGGTGCSACSAAAGATTGA-3') and either core200- (5'-TCYTCYTCYTCYTGCATIGC-3') or core157- (5'-CIGGRTCICCHATRTTIGC-3'). These primers correspond to amino acids VLGATKID, ANIGDP, and AMQEEEE, respectively (Vacquier *et al.*, 1995). (2) When complete 5' mature bindin sequences were not obtained during the first round of 5' RACE, new primers were designed at the 5' end of the obtained sequence; then a second round of RACE amplification was conducted. (3) A 5' preprobindin primer was designed based on a comparison of preprobindin sequences of *Moira clotho* (this study) to preprobindin sequences of *Arbacia* (Glabbe and Clark, 1991), *Strongylocentrotus* (Gao *et al.*, 1986; Minor *et al.*, 1991), and *Lytechinus* (Minor *et al.*, 1991). This primer, pro180 (5'-AAGMGIKCIAGYSCIMGIAAGGG-3'), which corresponds to the conserved amino acids KR(A/S)S(A/P)RKG of the preprobindin, was used in combination with exact primers from the bindin core to amplify mature bindin sequences 5' of the core from *Eucidaris tribuloides* testis cDNA. (4) Bindin sequences obtained from RACE were subsequently confirmed by amplification, cloning, and sequencing of full mature bindin sequences from testis cDNA.

Sequencing of both DNA strands was performed on an ABI 377 automated sequencer, and sequences were

edited using Sequencher 4.1 (Gene Codes Corp.). Sequences have been deposited in GenBank (Accession numbers AY126482-AY126485, AF530406). Published mature bindin sequences from a single exemplar from each of the five genera in which bindin had been previously sequenced were taken from GenBank. These representatives were *Strongylocentrotus purpuratus* (Accession number: M14487, Gao *et al.*, 1986), *Lytechinus variegatus* (M59489, Minor *et al.*, 1991), *Arbacia punctulata* (X54155, Glabe and Clark, 1991), *Echinometra oblonga* (U39503, Metz and Palumbi, 1996), and *Tripneustes ventricosus* (AF520222, Zigler and Lessios, 2003). Three amino acids of the core region of the bindin of *Lytechinus variegatus* [numbers 367 (N), 368 (L), and 385 (Y) in the alignment of Vacquier *et al.*, (1995)] were changed to A, V, and D, respectively, based on our own sequence data of *Lytechinus* bindin from 25 individuals representing 5 species; all 25 sequences had these amino acids at the 3 sites (Zigler and Lessios, unpub.). In *Echinometra oblonga*, sequences for the extreme 3' end of preprobindin are not in GenBank. They were inferred from the primer sequences used by Metz and Palumbi (1996) to amplify mature bindin sequences.

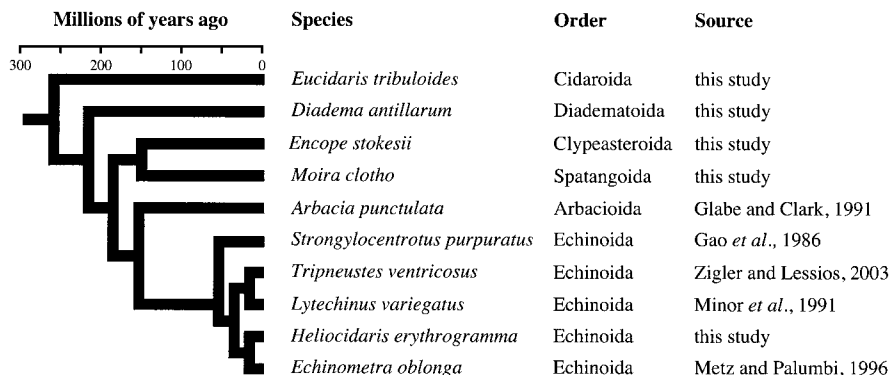
#### Sequence analysis

We aligned the mature bindin amino acid sequences with ClustalX ver. 1.81 (Thompson *et al.*, 1997), and adjusted the alignment by eye in Se-A1 (ver. 2.0a5, Rambaut, 1996). We characterized the amino acid changes observed in the core region of bindin as either radical or conservative with respect to charge and polarity (Taylor, 1986; Hughes *et al.*, 1990). The PROTPARAM tool of the EXPASY proteomics server of the Swiss Institute for Bioinformatics (<http://www.expasy.org>) was used to calculate Kyte and Doolittle (1982) hydrophobicity plots (window size = 11 amino acids) for each mature bindin sequence. The PROTSSCALE tool of the same server was used to calculate amino acid composition

for the mature bindins both for the core region (10 sequences, 55 amino acids per sequence) and for mature bindin sequences outside the core (10 sequences of varying length for a total of 1909 amino acids). The program CODONS (Lloyd and Sharp, 1992) was used to calculate the effective number of codons (ENC), a measure of codon usage bias (Wright, 1990), for each sequence. ENC values can range from 20 to 61, with 61 indicating that all synonymous codons are used in equal frequency (no codon bias), and 20 indicating that only a single codon is used for each amino acid (maximum codon bias). The statistical analysis of protein sequences (SAPS, [http://www.isrec.isb-sib.ch/software/SAPS\\_form.html](http://www.isrec.isb-sib.ch/software/SAPS_form.html)) program was used to identify separated repeats, simple tandem repeats, and periodic repeats in each mature bindin sequence (Brendel *et al.*, 1992).

### Results and Discussion

Figure 1 shows the phylogenetic relationships among the echinoid orders from which bindin was sequenced, as they have been reconstructed from molecular, morphological, and fossil evidence (Littlewood and Smith, 1995; Smith *et al.*, 1995). As the figure indicates, bindin is present not only in the Echinoidea and the Arbacioidea (from which it was previously known), but also in the sand dollars (Clypeasteroidea) and the heart urchins (Spatangoida), as well as the phylogenetically much more distant Diadematoidea and Cidaroida. Along with the sequence of *Heliocidaris*, reported in this paper, and the previously known sequences from *Arbacia*, *Strongylocentrotus*, *Tripneustes*, *Lytechinus*, and *Echinometra*, the data set covers orders that contain more than 70% of all extant echinoid species (Kier, 1977). The Cidaroida, the only extant order of the subclass Perischochinoidea, is the lineage most divergent from all other echinoids. It was separated from the Euechinoidea approximately 250 mya. Bindin's presence in both extant subclasses of the Echinoidea indicates that it was present in

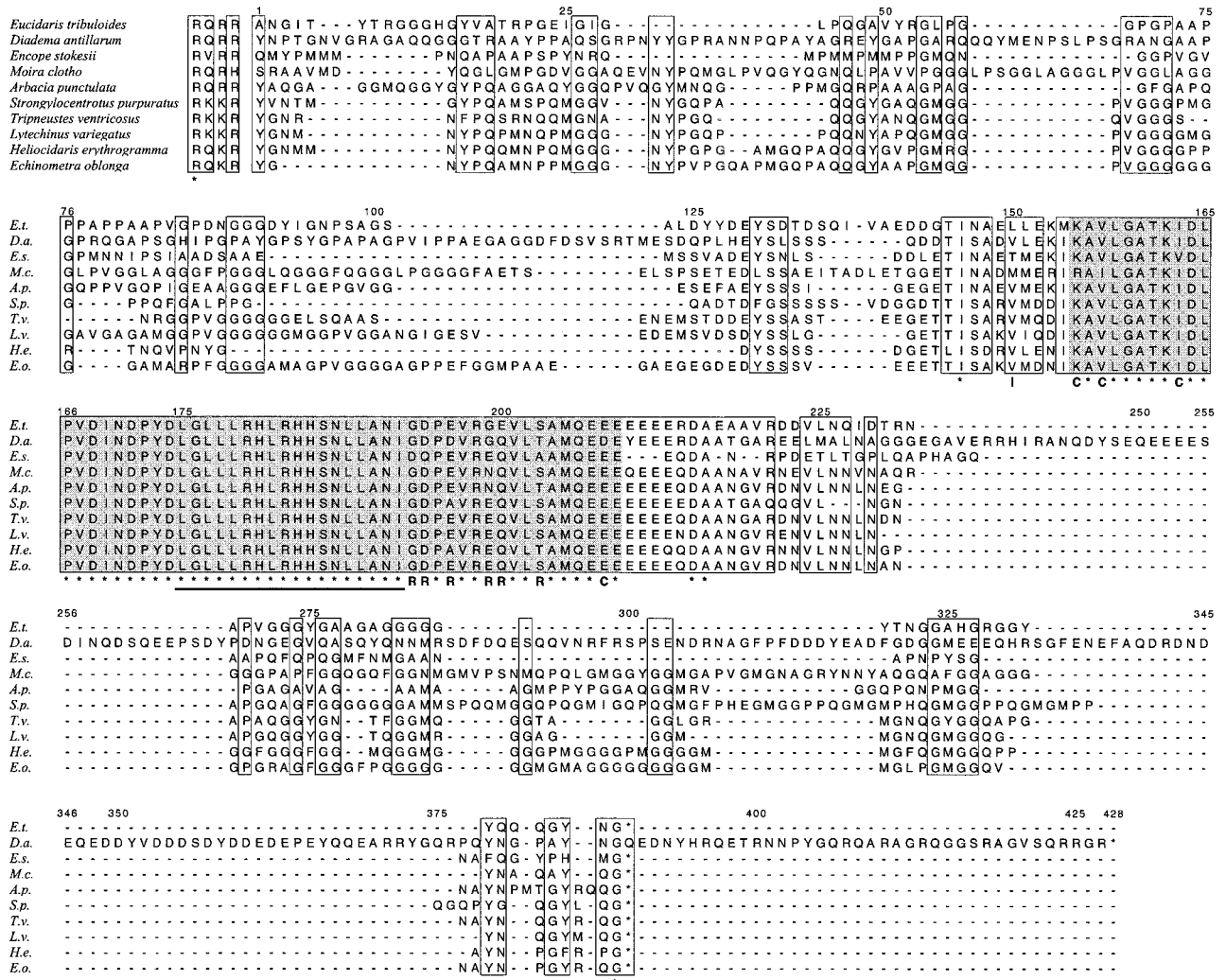


**Figure 1.** Phylogenetic relationships, divergence times, and systematic position of genera in which bindin has been sequenced. Echinoid phylogeny and divergence times are from Smith (1988) and Smith *et al.* (1995). Source of bindin sequence data is also indicated.

their common ancestor and that it has been evolving along each of the branches of the sea urchin phylogenetic tree for more than 250 my. Whether bindin is present in other echinoderms remains uncertain. Moy and Vacquier (1979) found that their antibody to *Strongylocentrotus purpuratus* bindin did not react with sperm from three species of sea stars, and “zoo blots” using *S. purpuratus* bindin sequences to probe genomic DNAs of a sea cucumber and a sea star were negative (Minor *et al.*, 1991). No attempt has been made to determine bindin’s presence in the ophiuroids or crinoids.

Figure 2 indicates that the aligned mature bindin sequences are a mosaic of highly conserved and highly diver-

gent regions. Over the past 250 my, the 55 residues of the core (amino acids 155–209) have been remarkably conserved. This region does not contain any insertions or deletions in any echinoid lineage. Of the 55 amino acids, 45 are conserved across all of the 10 exemplars, including a stretch of 29 residues in a row (amino acids 164–192). The B18 sequence of 18 amino acids implicated in membrane fusion (Ulrich *et al.*, 1998, 1999) is part of this perfectly conserved section. Seven amino acid sites in the core region exhibit a singleton amino acid change (*i.e.*, a change found in only one of the sequences). Four of these changes are conservative with respect to charge and polarity (amino



**Figure 2.** Mature bindin amino acid alignment. The first four amino acids are the presumed cleavage site from preprobindin. Sites at which amino acids are identical in more than 50% of the genera are enclosed in boxes. Conservation across all 10 genera is indicated by an asterisk below the site. Dashes indicate deletions. The site for which an intron is known to exist in *Echinometra*, *Arbacia*, *Strongylocentrotus*, *Lytechinus*, *Tripneustes*, *Heliocidaridaris*, and *Diadema* is indicated by an “I” under the alignment. The core region is shaded. The B18 region of the core is indicated by a bar beneath the alignment. Sites in the core where radical amino acid changes have occurred are marked with an ‘R’ under the alignment; sites in the core where only conservative amino acid changes have occurred are marked with a ‘C’. Stop codons are indicated by an asterisk after the last amino acid.

acids at positions 155, 157, 164, and 208), and three are radical (positions 193, 194, and 200). Each of positions 196, 199, and 203 contain three amino acids across the 10 genera, indicating that there have been at least two changes at each of these sites. At least one of the changes at each site must have been a radical change. Thus, radical changes are observed in only six amino acid positions of the core region, all of them concentrated in a small portion of the core close to the C terminus (amino acids 193, 194, 196, 199, 200, and 203). The rest of the core (amino acids 155 through 192 and 204 through 209) contains only four conservative singleton amino acid substitutions.

A second conserved region is the cleavage site at the border between preprobindin and mature bindin (Fig. 2). In *Strongylocentrotus purpuratus*, the cleavage site is marked by a motif of four basic amino acids (RKKR) (Gao *et al.*, 1986). Multibasic motifs are also present in the other nine genera (Fig. 2). Such multibasic motifs typically mark the cleavage sites of proproteins from the mature molecule during the secretory process through the action of proprotein convertases (Steiner, 1998; Seidah and Chretien, 1999). The conservation of this multibasic motif in bindin reinforces the idea that it functions as a signal for the cleavage of preprobindin from mature bindin in all echinoids.

In contrast to the core and to the cleavage site, the rest of the molecule is so variable between orders that we have little confidence that the alignment of these regions depicted in Figure 2 is correct. There is a great amount of variation in the length of mature bindin both on the 5' and on the 3' side of the molecule (Table 1). This study identifies both the longest and the shortest bindins described to date. Bindin in *Diadema antillarum* (418 amino acids) is more than twice as long as bindin in *Encope stokesii* (193 amino acids). Bindin length 5' of the core ranges from 78 to 148 amino acids, while bindin length 3' of the core ranges from 56 to 215 amino acids. There seems to be no discernible evolutionary trend in bindin length. Closely related orders do not

tend to have bindins that are of similar length. Indeed, it cannot be assumed that the genera we have included in the study are representative of their orders in this respect. The regions on either side of the core were found to confer species-specificity in *Strongylocentrotus* (Lopez *et al.*, 1993). If their variation reflects the requirements of this function, they can be expected to vary in a phylogenetically unpredictable fashion.

An intron located at a conserved position just 5' of the core region has been identified in the mature bindins of *Echinometra* (Metz and Palumbi, 1996), *Arbacia* (Metz *et al.*, 1998a), *Strongylocentrotus* (Biermann, 1998), and *Tripneustes* (Zigler and Lessios, 2003). In each of these genera, the intron is located at a conserved valine (amino acid 150 in Fig. 2). Comparison of sequences derived from both cDNA and from genomic DNA in *Heliocidaris* (Zigler *et al.*, 2003), *Lytechinus*, and *Diadema* revealed that in these genera the intron also exists at the same location and that its point of insertion is also a valine. We have made no attempt to amplify bindin from genomic DNA in *Eucidaris*, *Moira*, and *Encope*, so we do not know whether this intron is a universal feature of all bindins. These three genera do not have a valine in the site at which the intron is known to exist in the others, but the significance of this pattern cannot be evaluated with the present data.

Previous studies have identified both bindins with glycine-rich repeat structures and bindins that lack such structure. Glycine-rich repeats were found in the bindins of *Lytechinus* (Minor *et al.*, 1991), *Strongylocentrotus* (Biermann, 1998), *Echinometra* (Metz and Palumbi, 1996), and *Tripneustes* (Zigler and Lessios, 2003), all members of the order Echinoida. Consistent with the phylogenetic position of *Heliocidaris*, its bindin also contains glycine-rich repeat sequences, with MGGGN and VGGGGP on the 5' side of its core, and the series MGGG-MGGGGP-MGGGGP-MGGGGM-MGFQG-MGGQPP on the 3' side. Although *Moira* belongs to a different order, its bindin also contains extensive glycine-rich repeats, with the sequence PGGGL-PSGGL-AGGGL-PVGGL-AGGGL-PVGGL-AGGGF-PGGGL-QGGGF-QGGGL-PGGGG found 5' of the core. Glabe and Clark (1991) noted that bindin from *Arbacia punctulata* lacked significant repeat structure, and this observation was extended to three other species of *Arbacia* (Metz *et al.*, 1998a). *Eucidaris*, *Encope*, and *Diadema* resemble *Arbacia* in containing only minimal tandem or separated repeats, the longest of which is PAAP-PPAP-PAAP in the region flanking the 5' side of the core in *Eucidaris*. Thus, glycine-rich repeat structure remains a common trait of the bindin of the Echinoida, although, as the data from the spatangoid *Moira* indicate, it is not a characteristic limited to this order or even to a closely aligned clade.

There are no cysteine or tryptophan residues in any mature bindin. Disulfide bonds formed between cysteine residues are often critical for protein structure, and in rap-

**Table 1**

Number of amino acids in three regions of the mature bindin in 10 genera

	5'	Core	3'	Total
<i>Eucidaris</i>	101	55	60	216
<i>Diadema</i>	148	55	215	418
<i>Encope</i>	82	55	56	193
<i>Moira</i>	138	55	94	287
<i>Arbacia</i>	105	55	73	233
<i>Lytechinus</i>	103	55	60	218
<i>Tripneustes</i>	88	55	68	211
<i>Strongylocentrotus</i>	82	55	99	236
<i>Heliocidaris</i>	78	55	73	206
<i>Echinometra</i>	111	55	75	241

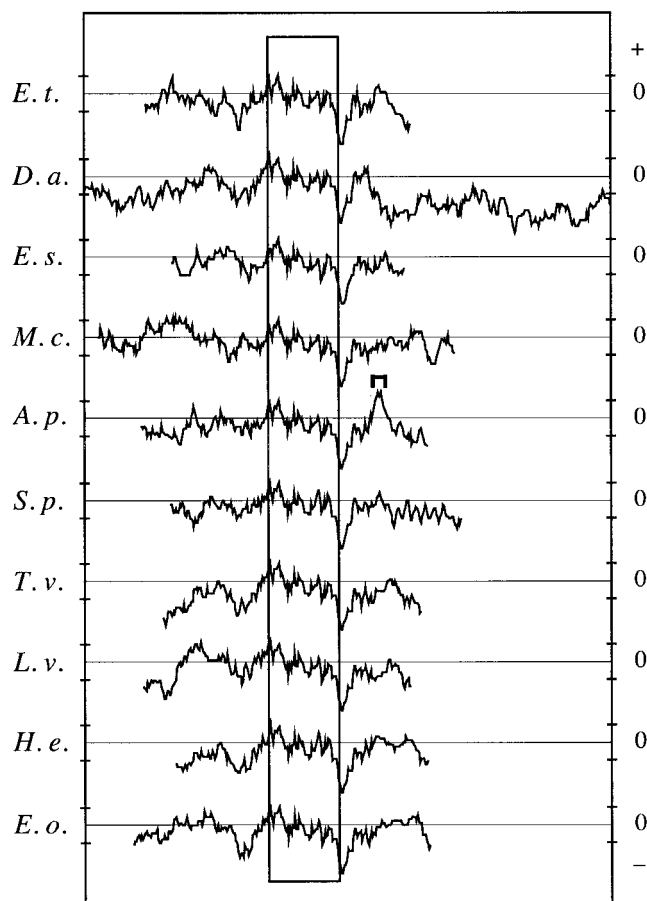
5' and 3' regions are defined relative to the conserved core.

idly evolving proteins—such as toxins of cone snails (Duda and Palumbi, 1999) and pheromones of the marine ciliate *Euplotes* (Luporini *et al.*, 1995)—cysteine residues are often among the most conserved amino acids, serving as guides for aligning sequences. Thus, the lack of cysteine residues in bindin may have important structural consequences. When all sequences are pooled, glycine is by far the most common amino acid outside the core, constituting nearly a quarter of all residues. If the orders that possess glycine-rich repeats (Echinoidea and Spatangoida) are separated from those that do not, glycine remains the most common amino acid in both categories, constituting 29.6% of the non-core amino acids in the former and 16.4% of non-core residues in the latter. The six most common residues outside the core (G, A, P, Q, N, and E) compose 63.9% of all non-core residues. Leucine is the most common amino acid in the core, present in 10 completely conserved amino acid positions, including 6 of the 18 amino acids in the B18 region. There is a much higher proportion of charged residues in the core (31.8%) than in the rest of the molecule (15.6%). Each of the five charged amino acids (E, D, R, H, and K) is more common in the core.

Another common feature of all bindins is their lack of codon usage bias. ENC values among the 10 genera range from 61 (for *Eucidaris* and *Diadema*) to 48.1 (for *Arbacia*), with an average of 56.4. Low levels of codon usage bias have also been observed in sex-related genes in *Drosophila* (Civetta and Singh, 1998) and in the *Chlamydomonas* mating-type locus genes *Mid* and *Fus1* (Ferris *et al.*, 2002).

Given the large divergence in amino acid sequence and length (and the uncertainties in alignments), it is not surprising that hydrophobicity plots (Fig. 3) from these bindins are diverse. The conserved amino acid sequence of the core and its flanking regions causes all plots to be similar through the middle of the molecule. Plots of the closely related *Tripneustes ventricosus*, *Lytechinus variegatus*, *Heliocidaris erythrogramma*, and *Echinometra oblonga* bindins are similar throughout their lengths. The rest of the hydrophobicity plots are not clearly similar. One particularly distinct region is the long hydrophilic stretches in *Diadema* bindin along its extended length. A second is the highly hydrophobic region 3' of the core of *Arbacia* bindin, noted by Glabe and Clark (1991).

The only other gamete recognition protein that has been studied in marine invertebrates separated for as long as 250 my is the gastropod sperm protein lysin. Lysin opens a hole in the vitelline envelope of free-spawning snails and thus enables sperm to penetrate to the plasma membrane of the egg. It has been studied in the abalones (*Haliotis*) (Lee and Vacquier, 1992; Lee *et al.*, 1995; Yang *et al.*, 2000; reviewed in Kresge *et al.*, 2001) and in two genera of turban snails, *Tegula* and *Norrisia* (Hellberg and Vacquier, 1999). Abalones and turban snails diverged 250 mya, roughly the same time the cidaroids separated from the euechinoids. The



**Figure 3.** Hydrophobicity plots of mature bindin in 10 genera. The genera are presented in the same order as in Figure 1 and abbreviated as follows: *E.t.*: *Eucidaris tribuloides*, *D.a.*: *Diadema antillarum*, *E.s.*: *Encope stokesii*, *M.c.*: *Moira clotho*, *A.p.*: *Arbacia punctulata*, *S.p.*: *Strongylocentrotus purpuratus*, *T.v.*: *Tripneustes ventricosus*, *L.v.*: *Lytechinus variegatus*, *H.e.*: *Heliocidaris erythrogramma*, and *E.o.*: *Echinometra oblonga*. A box encloses the core, and a bracket indicates the hydrophobic region in *Arbacia*. The scale bars above and below each zero line mark +1 and -1, respectively.

additional bindin sequences reported here reinforce the conclusions of Hellberg and Vacquier (1999) from comparisons between the modes of evolution of these two proteins. Although they are both involved in gamete recognition and both lack cysteine residues, they evolve in different fashions. There is no equivalent of a bindin core region in lysin; amino acid substitutions are spread throughout the molecule, with only three amino acids conserved between all *Haliotis* species and the two teguline genera. Instead of conserving a section of the molecule, lysin has maintained its function by conserving secondary structure through conservative amino acid substitutions (Hellberg and Vacquier, 1999). Length variation is another obvious difference. Mature bindin length varies from 193 to 418 amino acids, but lysin length (at least in the two groups studied to date) only from 126 to 138 amino acids.

## Conclusions

The comparisons of bindin from 10 genera of echinoids reveal the results of long-term evolution under two opposing selective forces acting on gamete recognition molecules. The sections of the molecule involved in the basic functions of gamete fusion and post-translational cleaving of the preprobindin have been remarkably conserved over 250 my of evolution, presumably through purifying selection. The sections involved in species recognition have been evolving rapidly in seemingly unpredictable directions, presumably under diversifying selection; such changes are likely to be specific to each species.

A number of features identified by these comparisons are in need of functional explanations. Among the conserved features, the lack of change in the core region is the only one that can be easily explained. We do not yet know whether there is a particular reason for the low codon usage bias of all bindins, for the absence of tryptophan or cysteine residues, or for the absence of major hydrophobic regions in all bindins except that of *Arbacia*. The differences between the orders are equally puzzling. Is there a functional reason for the length variation of the regions outside the core? Why do the Echinozoa and the Spatangoida have glycine-rich repeats in the regions flanking the core, while other orders do not? Comparisons alone cannot provide answers to these questions; but they can identify features of the molecule that are worthy of functional study.

## Acknowledgments

We are grateful to A. and L. Calderón for providing support in the laboratory, to M. McCartney for primer design and advice on the RACE technique, to T. Duda for collecting *Moiria clotho*, and to E. Popodi for providing testis RNA from *Heliocidaris erythrogramma*. Comments from C. Cunningham, D. McClay, R. Sponer, W. Swanson, V. Vacquier, and two anonymous reviewers improved the manuscript. This work was supported by National Science Foundation and Smithsonian predoctoral fellowships to KSZ, by the Duke University Department of Zoology, and by the Smithsonian Molecular Evolution Program.

## Literature Cited

- Biermann, C. H. 1998.** The molecular evolution of sperm bindin in six species of sea urchins (Echinozoa: Strongylocentrotidae). *Mol. Biol. Evol.* **15**: 1761–1771.
- Brandriff, B., G. W. Moy, and V. D. Vacquier. 1978.** Isolation of sperm bindin from the oyster (*Crassostrea gigas*). *Gamete Res.* **89**: 89–99.
- Brendel, V., P. Bucher, I. Nourbakhsh, B. E. Blaisdell, and S. Karlin. 1992.** Methods and algorithms for statistical analysis of protein sequences. *Proc. Natl. Acad. Sci. USA* **89**: 2002–2006.
- Civetta, A., and R. S. Singh. 1998.** Sex-related genes, directional sexual selection, and speciation. *Mol. Biol. Evol.* **15**: 901–909.
- Debenham, P., M. A. Brzezinski, and K. R. Foltz. 2000.** Evaluation of sequence variation and selection in the bindin locus of the red sea urchin, *Strongylocentrotus franciscanus*. *J. Mol. Evol.* **51**: 481–490.
- Duda, T. F., and S. R. Palumbi. 1999.** Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc. Natl. Acad. Sci. USA* **96**: 6820–6823.
- Ferris, P. J., E. V. Armbrust, and U. W. Goodenough. 2002.** Genetic structure of the mating-type locus of *Chlamydomonas reinhardtii*. *Genetics* **160**: 181–200.
- Gao, B., L. E. Klein, R. J. Britten, and E. H. Davidson. 1986.** Sequence of mRNA coding for bindin, a species-specific sperm protein required for fertilization. *Proc. Natl. Acad. Sci. USA* **83**: 8634–8638.
- Glabe, C. G., and D. Clark. 1991.** The sequence of the *Arbacia punctulata* bindin cDNA and implications for the structural basis of species-specific sperm adhesion and fertilization. *Dev. Biol.* **143**: 282–288.
- Hellberg, M. E., and V. D. Vacquier. 1999.** Rapid evolution of fertilization selectivity and lysin cDNA sequences in teguline gastropods. *Mol. Biol. Evol.* **16**: 839–848.
- Hofmann, A., and C. G. Glabe. 1994.** Bindin, a multifunctional sperm ligand and the evolution of new species. *Semin. Dev. Biol.* **5**: 233–242.
- Hughes, A. L., T. Ota, and M. Nei. 1990.** Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. *Mol. Biol. Evol.* **7**: 515–524.
- Kier, P. M. 1977.** The poor fossil record of the regular echinoid. *Paleobiology* **3**: 168–174.
- Kresge, N., V. D. Vacquier, and C. D. Stout. 2001.** Abalone lysin: the dissolving and evolving sperm protein. *Bioessays* **23**: 95–103.
- Kyte, J., and R. F. Doolittle. 1982.** A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* **157**: 105–132.
- Lee, Y.-H., and V. D. Vacquier. 1992.** The divergence of species-specific abalone sperm lysins is promoted by positive Darwinian selection. *Biol. Bull.* **182**: 97–104.
- Lee, Y.-H., T. Ota, and V. D. Vacquier. 1995.** Positive selection is a general phenomenon in the evolution of abalone sperm lysin. *Mol. Biol. Evol.* **12**: 231–238.
- Littlewood, D. T. J., and A. B. Smith. 1995.** A combined morphological and molecular phylogeny for sea urchins (Echinozoa: Echinodermata). *Philos. Trans. R. Soc. Lond. B* **347**: 213–234.
- Lloyd, A. T., and P. M. Sharp. 1992.** CODONS: A microcomputer program for codon usage analysis. *J. Hered.* **83**: 239–240.
- Lopez, A., S. J. Miraglia, and C. G. Glabe. 1993.** Structure/function analysis of the sea-urchin sperm adhesive protein bindin. *Dev. Biol.* **156**: 24–33.
- Luporini, P., A. Vallesi, C. Miceli, and R. A. Bradshaw. 1995.** Chemical signaling in ciliates. *J. Eukaryot. Microbiol.* **42**: 208–212.
- Metz, E. C., and S. R. Palumbi. 1996.** Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. *Mol. Biol. Evol.* **13**: 397–406.
- Metz, E. C., G. Gomez-Gutierrez, and V. D. Vacquier. 1998a.** Mitochondrial DNA and bindin gene sequence evolution among allopatric species of the sea urchin genus *Arbacia*. *Mol. Biol. Evol.* **15**: 185–195.
- Metz, E. C., R. Robles-Sikisaka, and V. D. Vacquier. 1998b.** Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitutions in introns and mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **95**: 10,676–10,681.
- Minor, J. E., D. R. Fromson, R. J. Britten, and E. H. Davidson. 1991.** Comparison of the bindin proteins of *Strongylocentrotus franciscanus*, *S. purpuratus*, and *Lytechinus variegatus*: sequences involved in the species specificity of fertilization. *Mol. Biol. Evol.* **8**: 781–795.
- Moy, G. W., and V. D. Vacquier. 1979.** Immunoperoxidase localization of bindin during the adhesion of sperm to sea urchin eggs. *Curr. Top. Dev. Biol.* **13**: 31–44.
- Rambaut, A. 1996.** Se-Al: Sequence Alignment Editor. University of

- Oxford, Oxford [Online]. Available: <http://evolve.zoo.ox.ac.uk/> [accessed June 2003].
- Seidah, N. G., and M. Chretien. 1999.** Proprotein and prohormone convertases: a family of subtilases generating diverse bioactive polypeptides. *Brain Res.* **848**: 45–62.
- Smith, A. B. 1984.** *Echinoïd Paleobiology*. George Allen and Unwin, London. 190 pp.
- Smith, A. B. 1988.** Phylogenetic relationship, divergence times, and rates of molecular evolution for camarodont sea urchins. *Mol. Biol. Evol.* **5**: 345–365.
- Smith, A. B., D. T. J. Littlewood, and G. A. Wray. 1995.** Comparing patterns of evolution: larval and adult life history stages and ribosomal RNA of post-Paleozoic echinoids. *Philos. Trans. R. Soc. Lond. B* **349**: 11–18.
- Steiner, D. F. 1998.** The proprotein convertases. *Curr. Opin. Chem. Biol.* **2**: 31–39.
- Swanson, W. J., and V. D. Vacquier. 2002.** Reproductive protein evolution. *Annu. Rev. Ecol. Syst.* **33**: 161–179.
- Takagi, T., A. Nakamura, R. Deguchi, and K. Kyojuka. 1994.** Isolation, characterization, and primary structure of three major proteins obtained from *Mytilus edulis* sperm. *J. Biochem.* **116**: 598–605.
- Taylor, W. R. 1986.** The classification of amino acid conservation. *J. Theor. Biol.* **119**: 205–218.
- Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997.** The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **24**: 4876–4882.
- Ulrich, A. S., M. Otter, C. G. Glabe, and D. Hoekstra. 1998.** Membrane fusion is induced by a distinct peptide sequence of the sea urchin fertilization protein bindin. *J. Biol. Chem.* **273**: 16,748–16,755.
- Ulrich, A. S., W. Tichelaar, G. Forster, O. Zschornig, S. Weinkauff, and H. W. Meyer. 1999.** Ultrastructural characterization of peptide-induced membrane fusion and peptide self-assembly in the lipid bilayer. *Biophys. J.* **77**: 829–841.
- Vacquier, V. D. 1998.** Evolution of gamete recognition proteins. *Science* **281**: 1995–1998.
- Vacquier, V. D., and G. W. Moy. 1977.** Isolation of bindin: the protein responsible for adhesion of sperm to sea urchin eggs. *Proc. Natl. Acad. Sci. USA* **74**: 2456–2460.
- Vacquier, V. D., W. J. Swanson, and M. E. Hellberg. 1995.** What have we learned about sea urchin sperm bindin? *Dev. Growth Differ.* **37**: 1–10.
- Wright, F. 1990.** The “effective number of codons” used in a gene. *Gene* **87**: 23–29.
- Yang, Z., W. J. Swanson, and V. D. Vacquier. 2000.** Maximum likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Mol. Biol. Evol.* **17**: 1446–1455.
- Zigler, K. S., and H. A. Lessios. 2003.** Evolution of bindin in the pantropical sea urchin *Tripneustes*: comparisons to bindin of other genera. *Mol. Biol. Evol.* **20**: 220–231.
- Zigler, K. S., E. C. Raff, E. Popodi, R. A. Raff, and H. A. Lessios. 2003.** Adaptive evolution of bindin is correlated with the shift to direct development in the genus *Heliocidaris*. *Evolution* (In press).